Screening for Depressed Individuals by Using Multimodal Social Media Data

Paulo Mann,[1] **Aline Paes,** [1] **Elton H. Matsushima** [2]
[1] Institute of Computing, Universidade Federal Fluminense, Brazil
[2] Department of Psychology, Universidade Federal Fluminense, Brazil
{paulomann@id, alinepaes@ic}.uff.br, eh.matsushima@gmail.com

Abstract

Depression has increased at alarming rates in the worldwide population. One alternative to finding depressed individuals is using social media data to train machine learning (ML) models to identify depressed cases automatically. Previous works have already relied on ML to solve this task with reasonably good F-measure scores. Still, several limitations prevent the full potential of these models. In this work, we show that the depression identification task through social media is better modeled as a Multiple Instance Learning (MIL) problem that can exploit the temporal dependencies between posts.

One of the most common mental disorders — depression — affects more than 300 million people across the globe (Organization et al. 2017). The statistics, however, do not always reflect reality *per se*, as many depressed individuals are kept unknown. Several reasons might contribute to this: individuals might not have the money, knowledge about the disorder, or they may fear social stigma to look out for help (Andrade, Alonso, and Mneimneh 2014), thus not being accounted for the statistics, and lacking adequate treatment.

An alternative to early detection at clinical attendance is to use ML models trained on annotated social media content to predict whether the person shows depression symptoms or not. Furthermore, these models can suggest which behavior, on social media, might lead to depression that differs from previously established psychiatric criteria — typically used on clinical consultations (Association et al. 2013). Such methodology have been proposed and explored by several previous studies, including ours (Mann, Paes, and Matsushima 2020; Reece and Danforth 2017; Shen et al. 2017; De Choudhury et al. 2013).

In our recently published work (Mann, Paes, and Matsushima 2020), we explored the role of each information modality posted in social media (IMAGE, TEXT, or IMAGE+TEXT) with neural network classifiers that were further compared to manually engineered features for the task of detecting depression. For that study, we have created the dataset with a sample of students' posts on Instagram (more information on the article (Mann, Paes, and Matsushima 2020)). We found that, in the best scenario, we can detect depression from both IMAGE+TEXT modalities with 0.79 of F-measure using deep feature extractors as ELMo (Peters et al. 2018) and ResNet34 for captions and images on Instagram. Although our learned model improves the performance of previously published research, they all share some limitations that could still hinder the full potential of the predictive model.

First, we argue that the nature of social media posts make them to be better formulated as a *Multiple Instance Learning* (MIL) task, which works on a weakly supervised learning regime (Carbonneau et al. 2018). In the MIL approach, data is arranged in sets (or bags), where instances contained in the bags are the minimum unit of examples, and the supervision is provided only for the entire set, and not for the instances (Carbonneau et al. 2018). Usually, for social media datasets, we have the user, which could be seen as a bag P^i, with k posts (instances) on the social media $P^i = \{p_1, ..., p_k\}$. For our dataset, we annotate the bags P^i (users) and not its respective instances (posts).

Previously published studies on the task of detecting depression using social media datasets either labeled the instances by just replicating the bag label into the instances of the bag, or by labeling posts satisfied by a string pattern (Reece and Danforth 2017; Shen et al. 2017; De Choudhury et al. 2013). Furthermore, they predict single instances p_k^i, and they usually do not report the results for predicting the bag P^i, as we did in our research by averaging the neural network output probabilities for all posts of each user, *i.e.,* the user-level scores.

Another observed particularity is the lack of explainability for deep learning models. Arguably, training specific models with manually engineered features allow for straightforward mechanisms for explainability. However, fine-tuning or domain adaptation of such models is limited compared to deep learning methodologies, particularly with unstructured data. Furthermore, fine-tuning is crucial for leveraging acquired knowledge to improve social media domains with low annotated resources.

To better understand the model's prediction, one crucial task is to find which information it is using to make decisions. By comprehending its internal choices, we become knowledgeable of its concepts and capture bugs in the learning process. To that, one recently published paper proposed a mechanism (called CHECKLIST) to evaluate the model's weaknesses and strengths by testing it against many types of

template sentences as input (Ribeiro et al. 2020). By doing that, they evaluate several capabilities of the model, ranging from Named Entity Recognition (NER) to vocabulary and robustness to noise. However, the proposed methodology is too general and does not assess the psycholinguistics particularities of our task. Moreover, they do not compare the impact of different fine-tuning strategies when confronted with these tests. Thus, we envision developing a new suite of tests explicitly designed to evaluate the depression detection task alongside general purpose tests, which we call DEPRESSION CHECKLIST.

Lastly, another observed limitation is the way that the examples are provided to the model. For the case of detecting depression, most of the previous studies used posts p_k^i as the input to feed the model disregarding their temporal order. However, temporal dependency between posts might be crucial for detecting depression with social media data. Taking the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) as inspiration, psychiatrists are encouraged to reason about the last two weeks of symptoms in the patient to make the final diagnosis (Association et al. 2013). This means that, potentially, we could enrich neural representations by leveraging the hierarchical or sequential dependency between posts in the set P^i. Note that the size of the set P^i can vary a lot among users due to differences in social media usage, and this information in itself might be important to take into account.

Research Goals and Outcomes We intend to develop methods to address the limitations of current methodologies by using a Transformer architecture (Vaswani, Shazeer, and Parmar 2017) and testing the models with the DEPRESSION CHECKLIST suite of tests. Moreover, we aim at demonstrating that our proposed solution can be generalized to other MIL and multimodal tasks.

By using the attention mechanism, Transformer is able to model the dependency between elements of a set. In the MIL setting, we can feed the model with our set of posts P^i, in which the model will find a relationship between the instances p_k of the bag. In that way, the initial matrices Q, K, and V are the *post representations* of the bag (packed together) as in Equation 1, with $d_k = dim(K)$.

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \qquad (1)$$

To allow for the multimodality scenario, we can use a deep feature extractor to retrieve each embedding and feed the *encoder* with textual (or visual) embeddings, whereas the other modality is used as input to the *decoder* network. With that, we expect that the *encoder-decoder* attention layers learn good representations based on both hierarchical (or sequential) dependencies *and* multimodality of posts at the same time. The model jointly learns the attention between the outputs K (keys) and V (values) of the *encoder* with the output Q (query) of the previous decoder layer at the *encoder-decoder* attention layer, which could be seen as a cross-modality layer. By leveraging the power of this architecture, we can model temporal dependencies, multimodality, and extract some form of restricted explanation by using the attention scores. Note that this idea generalizes to other MIL problems with two modalities that also benefit from the hierarchical dependency aspect, and that we would like to experiment with other publicly available datasets, as in (Kruk, Lubin, and Sikka 2019).

Timetable and Expected Contributions This project intends to contribute with methods to improve individuals' social welfare by early screening of depression symptoms over social media with AI. Potentially, the new methods developed here can also be employed to solve other problems with similar characteristics, namely: (1) multimodal data gathered from social media; (2) set of instances to compose a single example; (3) and temporal dependency over examples. The student and the advisor have already contributed with a first approach to partially handle those aspects with promising results on real-world data (Mann, Paes, and Matsushima 2020)[1]. For future directions, we expect to finish all DEPRESSION CHECKLIST experiments and write a proper article by February 2021. Following the paper submission, we will start working in the Transformer network.

References

Andrade, L. H.; Alonso, J.; and Mneimneh, Z. 2014. Barriers to mental health treatment: results from the WHO World Mental Health surveys. *Psychological medicine* 44(6): 1303–1317.

Association, A. P.; et al. 2013. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Carbonneau, M.; et al. 2018. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition* 77: 329–353.

De Choudhury, M.; Gamon, M.; Counts, S.; and Horvitz, E. 2013. Predicting depression via social media. *ICWSM* 13: 1–10.

Kruk, J.; Lubin, J.; and Sikka, K. 2019. Integrating Text and Image: Determining Multimodal Document Intent in Instagram Posts. In *EMNLP-IJCNLP*, 4622–4632. ACL.

Mann, P.; Paes, A.; and Matsushima, E. H. 2020. See and Read: Detecting Depression Symptoms in Higher Education Students Using Multimodal Social Media Data. In *ICWSM*, volume 14, 440–451.

Organization, W. H.; et al. 2017. Depression and other common mental disorders: global health estimates. Technical report, WHO.

Peters, M.; et al. 2018. Deep Contextualized Word Representations. In *Proc. of the 2018 Conf. of the NAACL*, 2227–2237. AC.

Reece, A. G.; and Danforth, C. M. 2017. Instagram photos reveal predictive markers of depression. *EPJ Data Science* 6(1): 15.

Ribeiro, M. T.; et al. 2020. Beyond Accuracy: Behavioral Testing of NLP Models with CheckList. In *ACL*, 4902–4912. Online: ACL.

Shen, G.; et al. 2017. Depression detection via harvesting social media: A multimodal dictionary learning solution. In *IJCAI*, 3838–3844.

Vaswani, A.; Shazeer, N.; and Parmar, N. 2017. Attention is all you need. In *NeurIPS*, 5998–6008.

[1]The research was conducted under the approval of the ethical committee of the University, CAAE: 89859418.1.0000.5243.